Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 41 (2012) 319 – 325

Engineering Procedia

www.elsevier.com/locate/procedia

International Symposium on Robotics and Intelligent Sensors 2012 (IRIS 2012)

Biologically Inspired Temporal Sequence Learning

Nooraini Yusoff[a]*, André Grüning[b]

[a]School of Computing,UUM College of Arts and Sciences, Universiti Utara Malaysia, 06010 UUM Sintok, Kedah, Malaysia
[b]Department of Computing, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford GU2 7XH, Surrey, UK

Abstract

We propose a temporal sequence learning model in spiking neural networks consisting of Izhikevich spiking neurons. In our reward-based learning model, we train a network to associate two stimuli with temporal delay and a target response. Learning rule is dependent on reward signals that modulate the weight changes derived from spike-timing dependent plasticity (STDP) function. The dynamic properties of our model can be attributed to the sparse and recurrent connectivity, synaptic transmission delays, background activity and inter-stimulus interval (ISI). We have tested the learning in visual recognition task, and temporal AND and XOR problems. The network can be trained to associate a stimulus pair with its target response and to discriminate the temporal sequence of the stimulus presentation.

Keywords: Temporal sequence learning; Spiking neural networks; Spike-timing dependent plasticity; Reward-based learning.

1. Introduction

Stimuli can be associated through many ways including sharing of properties, semantic relevance, and sequence correlation. Following the causality law of the world, in which an event precedes an effect, many causal relationships are temporally related. This also relates for sensor-sensor, sensor-motor and motor-motor events [1]; yellow before green for traffic lights (sensor-sensor), a visual image triggers utterance (sensor-motor), and an action of grasping after arm movement (motor-motor). These events are temporally correlated. Furthermore, several findings from behavioural experiments (e.g. [2], [3]) have found the effect of priming in visual recognition that shows signs of influence of previous information on the perception of subsequent information [4]-[6]. The effect is a result of 'spread activation' mechanism in the brain in which a recently probed stimulus invokes its associated information, consequently facilitating the retrieval of information of a later proceeded stimulus when both are related. For this case, the prime stimulus acts as a cue for the later stimulus.

On the other hand, understanding a cognitive process for its functional and structural behaviour is rather difficult as the brain is complex and there are times exhibiting chaotic patterns [7]. The consistency of findings between studies at the cellular level from the neuroscience field and psychological behaviour experiments still remains an intriguing subject to unearth. However, interesting progress has been made in artificial neural network research as a result of deeper understanding of the brain giving more meaningful biological interpretation to a model.

The findings from neurophysiological experiments provide clues how information is encoded in the brain. Biological neurons use pulses or spikes to transmit information across brain regions. It is now well accepted that computational significance lies in the timing of those spikes [8], [9]. Therefore, with their biological counterparts in the use of spikes for neuron communication and computation, the niche of the recent generation neural network models is ascribed to their

* Corresponding author. Tel.: +6-04-9284620; fax: +6-04-9284753 .
E-mail address: nooraini@uum.edu.my

1877-7058 © 2012 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
doi:10.1016/j.proeng.2012.07.179

spatio-temporal information encoding. Spatio-temporal neural networks are commonly known as spiking neural networks (SNNs). Nevertheless, the most challenging task in SNN models is neuronal activity encoding. How the activity dynamics could be efficiently analysed with respect to some noisy stimulus signals without losing the most of essential information [10].

From past studies of SNN, we have found that little work has been reported on its implementation in reinforcement learning paradigm. There are not many algorithms developed with explicit neural modelling. Only recently there seems to be increasing work on modelling of reinforcement learning in SNN, after neurophysiological data linking dopamine signals in the brain that is believed to play an important role in enhancement of synaptic changes [12], [13]. The dopamine signals are hypothesised to be responsible for the reward acquisition mechanism in the brain, thus giving us some clue on connection between synaptic plasticity at the microscopic level with behavioural changes in animals.

In reinforcement learning, agents must update their internal parameters in order to maximise reward over time at a given task [1], [10], [11], [14]. This is implemented through a series of trial-and-error action-rewards in response to environmental stimuli. Unlike supervised and unsupervised approaches, where in most cases learning follows some specific rules with given initial state, in the reinforcement approach, agents explore and exploit their unknown identity states to establish a learning. In agreement with the findings relating the role of dopamine signals in the brain reward circuits where at the cellular level they consolidate synaptic plasticity process, this type of learning is more dynamic and natural.

In this paper, we propose a stimulus-stimulus association learning scheme for spiking neural networks. In a stochastic spiking network with Izhikevich neurons [15], an agent is trained to associate delayed paired stimuli. We hypothesise that by priming the agent with a cue stimulus could facilitate the response to a later stimulus. The recurrent connectivity, properties of regular and fast spiking neurons, and synaptic transmission delays enforced in the proposed network, adapted from [16], provide richer dynamics to learning.

2. Simulation Model

The network model consists of excitatory neurons (80%) and inhibitory neurons (20%), the ratio of pyramidal cells (i.e. excitatory) to interneurons (i.e. inhibitory) in the cortical network [18], [19], with sparse and random connectivity (no self-feedback). Each excitatory neuron is randomly connected to 100 neurons, and each inhibitory neuron is randomly connected to 100 excitatory neurons only, with synaptic transmission delays between 1 to 20 ms [16]. Neurons are divided into subpopulations of stimulus groups (*S)*, response groups (*R)*, non-selective neurons (*NS)* and inhibitory pool (*IH*). A stimulus group is composed of a number of excitatory neurons that are selective to a stimulus. Meanwhile a response group consists of both excitatory and inhibitory neurons. For lateral inhibition to competitor group(s), each excitatory neuron in the response group has random connections to neurons from its inhibitory pool. The inhibitory pool is connected randomly to its competitor's excitatory neurons. Therefore, triggering a response group would invoke its inhibitory neurons which consequently to prevent activation of its competitors. Neurons from the *NS* group are not selective to any stimulus but their activity also contribute to learning dynamics, and *IH* consists of inhibitory neurons (ones that are not selected for lateral inhibition), see Fig. 1.

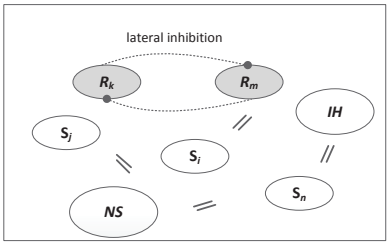

Fig. 1. Recurrent spiking network with subpopulations of stimulus groups (*S*), response groups (*R*), non-selective neurons (*NS*) and inhibitory pool (*IH*). Lines end with solid circle highlight inhibitory connections between response groups. (Please see the text above for details).

The spiking dynamics for each neuron are based on Izhikevich spiking neuron model [15] as in 1-3.

$$v' = 0.04v^2 + 5v + 140 - u + I \quad (1)$$

$$u' = a(bv - u) \quad (2)$$

$$if\ v \geq +30\ mV,\ then\ u \leftarrow u + d,\ v \leftarrow c \quad (3)$$

where v is the membrane potential, u is the recovery variable and I is the input, while a-d are the model parameters. The parameters a and b affect the recovery variable u, respectively is the time scale of u and the sensitivity of u to the sub-threshold fluctuations of v. The parameter c is the after-spike reset value of the membrane potential v, and d is the after-spike reset of the recovery variable u. If v reaches its peak at 30 mV, the membrane potential is reset to its resting state, c = -65 mV. To simulate the properties of real cortical neurons, all excitatory neurons exhibit regular spiking type neurons and all inhibitory neurons are fast spiking neurons [16].

3. Learning Implementation

In every learning simulation, for the first 100 ms, we initiate a network with random activity. For this purpose, we stimulate an arbitrary neuron with 20-pA current for every ms. With the same random activity as the background, the network is given a set of pair-response mappings $(S_i,S_j) \rightarrow R_k$, with different pairing strategies depending on the task. For each learning trial, at time t_n we present the first stimulus, i.e S_i to the network by stimulating all neurons in S_i with a strong current of 20 pA. After an inter-stimulus interval (ISI), we stimulate all neurons with the same amount of current to the second stimulus, i.e. S_j to be associated to S_i. An optimal ISI is chosen from a range of 10 – 50 ms based on a preliminary experiment. Each learning runs for 20 minutes simulated time with random presentation of stimulus pairs.

Within a 20-ms time window from the onset of the second stimulus, we count the number of activations in the response groups, i.e, R_k. The response group with the highest number of activations is considered to be the winner. To accelerate the learning, some bias current is supplied to the target winner. This is implemented via stimulation of 20-pA current to arbitrary neurons (with probability of neurons to be selected is between p=0.25 to 0.5, weak to strong potentiation) in the target response group. The next learning pair is presented after a 100-ms delay from the offset of each response interval. Synapse reinforcement is implemented based on a reward policy. The network is positively rewarded if a target response group is the winner of a learning trial, or otherwise negatively rewarded. The reward policy determines the amount of synapse potentiation (i.e. strong or weak potentiation) or depression.

During the 20-ms response interval, for the first 10 ms, we reward the network based on the number of activations in the response inhibitory groups. This is a way to reinforce the synapses for connectivity between a stimulus and the target response inhibitory group for preventing the activation of response competitor groups. Then we reward the network for the number of activations in the response excitatory groups within the 20-ms response interval for synapse reinforcement from the stimulus group to the target response excitatory group. The bias current is supplied to both winners of the target response inhibitory and excitatory groups. An example of spike raster plot at the early phase of learning and after a number of trials is depicted in Fig. 2.

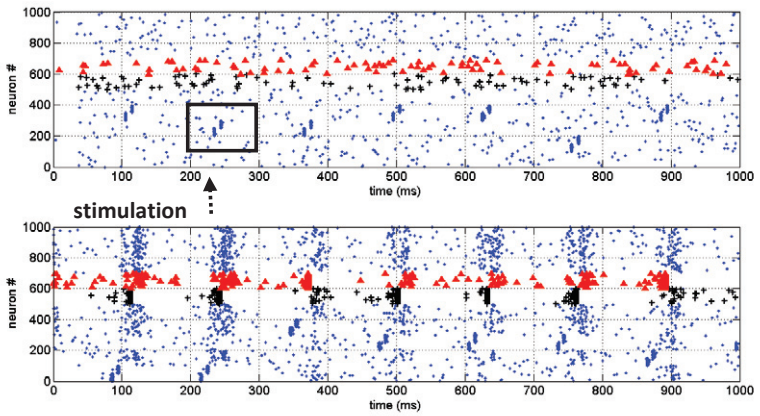

Fig. 2. Spike raster plot of a learning (*top*) at the early phase, and (*bottom*) after 500 seconds, within 1000 ms time window. Neurons in the response groups are marked with '+' (neurons 501-600) and '^' (neurons 601-700).

With the same network and stimulus presentation settings, we implement a series of probe trials (i.e. testing) to recall learned stimulus pairs (see Fig. 3). Furthermore, to see the recall performance with noisy stimuli, we also vary the number of neurons for random stimulation in learned groups with probability of less than 1.0. The testing result shows the averaged percentage of performance over a number of trials, i.e. *performance = (number of correct recall/number of trials)*100.*

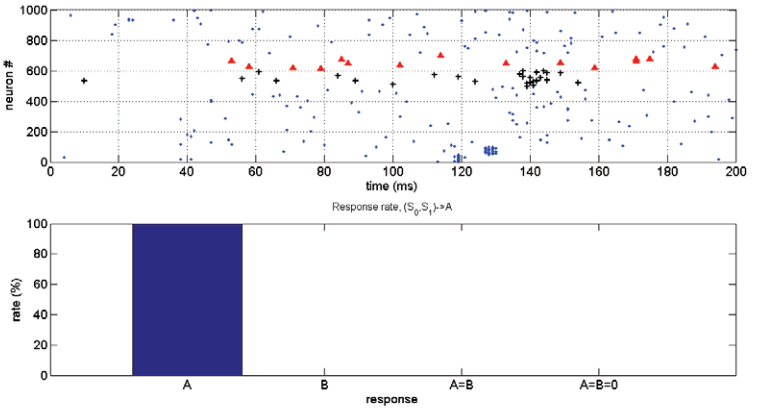

Fig. 3. One of probe trials for $(S_0,S_1) \rightarrow R_A$ within 200 ms time window. At random time t, between 100 to 120 ms, stimulus S_0 (neurons 1-50) is intensified with super threshold current 20 pA followed by stimulus S_1 (neurons 51-60) after an ISI. (*top*) Correct response recall of a trial where number of spikes from target response group R_A (neurons 501-600) is greater than number of spikes in R_B (neurons 601-700), within 20-ms response interval after the onset of S_1. (*bottom*) The averaged percentage of number of correct recalls over a number of probe trials for $(S_0,S_1) \rightarrow R_A$.

In our model, synaptic plasticity is implemented on excitatory synapses only for every 10-ms time step. The synaptic efficacy is dependent on a reinforcement signal (i.e. reward signal), $r(t)$, derived from a reward policy. The signal modulates the synaptic changes read from a spike-timing dependent plasticity (STDP) function (as in 4).

$$\Delta w_{stdp} = \Theta\{A_+ e^{-\Delta t/\tau_+},\ \Delta t \geq 0;\ A_- e^{\Delta t/\tau_-},\ \Delta t < 0\} \qquad (4)$$

From 4, the synapse is potentiated if the difference in firing times (Δt) between a postsynaptic neuron and its presynaptic neuron (i.e. t_{post}-t_{pre}) is ≥ 0, otherwise the synapse is depreciated. The magnitude of potentiatiaon (depression) is given by $A_+ e^{-\Delta t/\tau_+}$ ($A_- e^{\Delta t/\tau_-}$), where A represents the maximal change when the spike timing difference Δt is approaching 0, and τ is the time constant (in ms). For our STDP curve, $\tau_+ = \tau_- = 20$ ms, $A_+ = 0.1$, and $A_- = 0.15$ (following [17]).

The reinforcement signal $r(t)$ is obtained from a reward policy that is based on the number of neuron firings (F) of response groups within a response interval of 20 ms. The reward policy to derive $r(t)$ is given by:

$$r(t) = \begin{cases} r(t-1) + 0.5 & \text{if } F_i \geq 2F_j \quad (\textit{strong +ve reward}) \\ 1 - F_j/F_i & \text{if } F_j < F_i < 2F_j \quad (\textit{weak +ve reward}) \\ -0.1 & \text{if } F_i < F_j \quad (\textit{-ve reward}) \end{cases} \qquad (5)$$

where F_i and F_j are the number of firings of a target response group, and non-target group, respectively. The reinforcement signal rate is computed based on the type of reward; strong positive, weak positive and negative reward. The signal determines the amount of modulation to the summation of Δw_{stdp}. Therefore, the reward modulated STDP learning holds [14], [17]:

$$\Delta w(t) = [\alpha + r(t)]\, z(t) \qquad (6)$$

where α is the activity-independent increase of synaptic weight, *r(t)* and *z(t)* are the reinforcement signal (5) and the eligibility trace, respectively. *z(t)* represents the summation of Δw_{stdp} obtained from (4). Excitatory and inhibitory weights are initialised to 1.0 and -1.0, respectively. To avoid infinite saturation, weights are kept to be in the range between 0 and 4 mV.

4. Simulation Results

We ran a series of learning simulations with different pairing strategies. We varied the temporal sequences in learning with both exclusive and non-exclusive stimulus pairs. For the former case, learning was implemented in a visual recognition task. The neural network was trained to associate two visual stimuli to a response. For this task, learning only involved non-overlapping stimuli for different pairs. For learning with non-exclusive stimulus pairs, the neural network was trained to detect temporal sequences in tasks that simulated the AND and XOR problems.

4.1. Visual recognition task

For training with real visual association task, the network was rewarded for responses that pointed to correct match of paired visual images. Those visual images were preprocessed using Matlab Image Processing Toolbox. For simplicity and to reduce the amount of data required, each image was converted from RGB to grayscale and compressed to 20x20 pixels. There were 6 visual stimulus groups, i.e. *IMG = {apple, book, orange, pen, sun, sunglasses}*, consisting of exclusive excitatory neurons with 400 neurons each (i.e. 400 pixels = 400 neurons). From the Matlab preprocessed images, we further discretised their pixels (in the range of 0 to 1) into 0 and 1 with threshold of 0.5. For training purposes, pixels were represented by an array of 0s and 1s that each bit of 1 would invoke 20-pA 1-ms pulse current to a neuron. In our simulation, we designed a network consisting of 4000 spiking neurons with 3200 (80%) excitatory and 800 (20%) inhibitory neurons. Each excitatory neuron was randomly connected to 100 neurons, meanwhile each inhibitory neuron was connected to only 100 excitatory neurons.

The network was trained to associate a pair of images with a target response, R_A or R_B. A response group consisted of 400 excitatory neurons. The learning pairs were as follows: *{(apple, orange)* $\rightarrow R_A$*, (book, pen)* $\rightarrow R_B$*, (sun, sunglasses)* $\rightarrow R_A$*}*. Each learning simulation was run for 20 minutes, with approximately 2940 trials for each pair. The averaged learning performance was 89.46% and all image pairs were correctly discerned with 100% accuracy in probe trials. In addition, we also probed the network with distorted images. For distortion method, for each learned image pattern, a random pattern was created with probability of each neuron to be stimulated was 0.25. A distorted image pattern was derived from an XOR operation between the learned image pattern and the randomly generated pattern. For probe trials with distorted images (with 10 distorted versions for each image), the averaged correct response rate was achieved at 87.33% (see Table 1). An example of network activity during learning is depicted in Fig. 4.

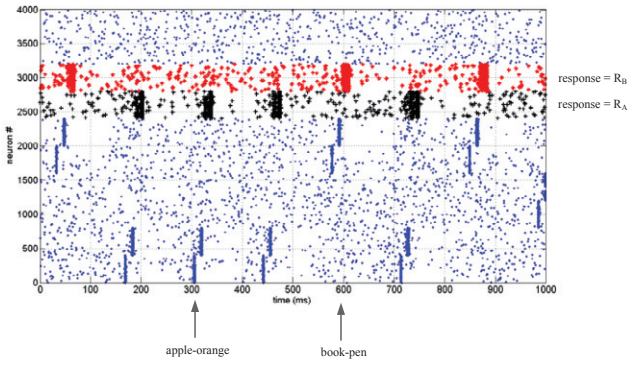

Fig. 4. Spike raster plot of network activity during learning after a number of trials. Each visual stimulus pair is reinforced to respond to a target group, i.e. R_A (neurons 2401-2800) or R_B (neurons 2801-3200). The number of spike counts in the response group within a 20-ms response interval determines the winning response.

Table 1. The recall results for training and testing for pair-response = *{(apple, orange) $\rightarrow R_A$, (book, pen) $\rightarrow R_B$, (sun, sunglasses) $\rightarrow R_A$}*

Image Pair	Target	Training (%)			Testing (with learned stimuli) (%)			Testing (with distorted stimuli) (%)		
		R_A	R_B	R_A=R_B	R_A	R_B	R_A=R_B	R_A	R_B	R_A=R_B
(apple, orange)	R_A	**95.78**	2.99	1.24	**100**	-	-	**100**	-	-
(book, pen)	R_B	7.49	**91.75**	0.76	-	**100**	-	15.5	**78.50**	6
(sun, sunglasses)	R_A	**80.86**	16.92	2.22	**100**	-	-	**83.50**	14.5	2

* R_A=R_B indicates no answer, i.e. the number of activations in group A and B is equal; percentage of correct recalls is highlighted in bold

4.2. Learning temporal logics: AND and XOR problems

In the previous experiment, the connectivity between neurons in a trained network was sparse and random with $p = 0.1$. For learning with exclusive stimulus groups, with a simple network structure, the proposed rule achieved excellent performance as all stimulus pairs were stable. However, the simplicity of the structure has some limitations for learning with high competition. The stability of each stimulus pair is dependent on correlation of spike patterns between two learning pairs. Furthermore, the connectivity between output neurons might also cause undesired causal firings. For example, for learning with 2 competing target responses, R_A and R_B, in which strengthening of synaptic strength between $S_i \rightarrow R_A$ could also lead to activation of neurons in response group R_B due to triggering of synapses $R_A \rightarrow R_B$. In short, firings of postsynaptic neurons of R_A in R_B. Therefore, there is a need to apply an inhibition mechanism via some anatomical constraints. To improve the discrimination rate in a more competitive learning, we suggest a modified network topology with a lateral inhibition mechanism (as illustrated in Fig. 1).

In the following experiment, lateral inhibition is applied for learning with non-exclusive stimulus groups. From Fig.1, in a network with 1000 neurons consisting of 800 excitatory and 200 inhibitory neurons, each excitatory neuron is connected to 100 neurons (excitatory and inhibitory), whilst each inhibitory neuron has connections to 100 excitatory neurons only. The number of excitatory neurons in each stimulus group is 50, and there are 150 neurons in a response group composed of 100 excitatory and 50 inhibitory neurons. To reduce too high correlation in spike patterns, for the same stimulus that exists in more than one pairs, we allow some degree of non-overlapping neurons for the stimulus groups with the same label.

The network with lateral inhibition was experimented in conditions with temporal logic problems, XOR and AND. For this purpose, we defined 4 distinct stimulus groups, S_0, S_1, S_2, and S_3. Here S_0 = TRUE for the first stimulus, and S_2 = TRUE for the second stimulus, and, S_1 and S_3 represented the FALSE value of the first and second stimuli, respectively. Meanwhile, the response group R_A represented a TRUE response and the response group R_B was considered a FALSE response. Therefore, for the AND problem, the pair-response set was $\{(S_0,S_2) \rightarrow R_A, (S_0,S_3) \rightarrow R_B, (S_1,S_2) \rightarrow R_B, (S_1,S_3) \rightarrow R_B\}$, and for the XOR problem pair-response set was = $\{(S_0,S_2) \rightarrow R_B, (S_0,S_3) \rightarrow R_A, (S_1,S_2) \rightarrow R_A, (S_1,S_3) \rightarrow R_B\}$,

Simulation result indicated that a network with stochastic dynamics and minimal anatomical constraints can also learn temporal logic functions with good performances achieved at 85.95% in training and 85.85% in testing for AND problem, and for learning temporal XOR at 81.88% in training and 79.53% in testing.

5. Conclusion

We propose temporal sequence detection via reward-based learning in dynamic environment setting. Learning is implemented in a stochastic way in which the network with random activity does not have any prior knowledge regarding the identity of learning signals. Input stimulation is induced at certain time only through perturbation to the network activity. Unlike in other goal-directed learning approaches that require so called 'teacher signals', in our model, learning is only dependent on global reinforcement signal (inspired by work in [17]). The signal consolidates synaptic changes by STDP process. In our experiments, at the early phase of learning, in addition to the background activity, the activation of neurons was only due to coincident firings triggered by the stimulated groups. After a number of rewards given based on the activation rate of the response groups, the network response to the paired stimuli became reinforced. The synaptic connections from the paired stimuli were stronger compared from other non-reinforced stimulus groups. Hence, neurons in paired stimuli could strongly influence their postsynaptic targets.

The learning model has been successfully tested for temporal sequence learning with exclusive stimulus groups as well as in a setting with overlap of patterns between stimulus groups. For the latter case, the model performance has been verified in solving temporal AND and XOR problems. In learning with non-exclusive stimulus groups, greater influence from the first stimulus is required to facilitate discrimination of target responses due to correlation in spike patterns. The optimal ISI for such learning condition has been found at 10 ms. For learning stability, we run every simulation for 20 minutes simulated time (1200 secs). However, we have observed that the learning gradient only deviates essentially from 0 within the first 10 secs, approximately just after 74 trials with random presentation of learning pairs.

Our work extends the reinforcement learning proposed in [17] by training a network to learn temporal sequences using lateral inhibition mechanism. With richer properties of network dynamics, we as well implement random input-output mappings, i.e. stimulus-pair $\rightarrow$ response. For the work in progress, we are extending the learning model to detect temporal sequence consisting of more than 2 stimuli. This could lead to some potential applications such as robot path tracking, linguistic processing and source localisation system.

Acknowledgements

This research has been funded from Ministry of Higher Education (Malaysia), and supported by the Engineering and Physical Sciences Research Council (EPSRC) grant (EP/I027831/1).

References

[1] Worgotter, F., Porr B., 2005. Temporal Sequence Learning, Prediction, and Control: A Review of Different Models and Their Relation to Biological Mechanisms, Neural Computation 17, p. 245.
[2] Erickson, C. A., Desimone, R., 1999. Responses of Macaque Perirhinal Neurons during and after Visual Stimulus Association Learning, Journal of Neuroscience 19(23), p. 10404.
[3] Naya, Y., Yoshida, M., Miyashita, Y., 2003. Forward processing of long term associative memory in monkey inferotemporal cortex, J. Neurosci. 23, p. 2861.
[4] Tulving, E., Schacter, D. L., Stark, H. A., 1982. Priming Effects in Word Fragment Completion are independent of Recognition Memory, Journal of Experimental Psychology: Learning, Memory and Cognition 8(4), p. 336.
[5] Schacter, D. L., 1992. Priming and multiple memory systems: Perceptual mechanisms of implicit memory, Journal of Cognitive Neuroscience 4(3), p. 244.
[6] Filippova, M. G., 2011. Does Unconscious Information Affect Cognitive Activity?: A Study Using Experimental Priming, The Spanish Journal of Psychology 14(1), p. 20.
[7] Korn H., Faure P., 2003. Is there chaos in the brain? II. Experimental evidence and related models, C R Biol. 326(9), p. 787.
[8] Maass, W., 1997. Networks of spiking neurons: The third generation of neural network models, Neural Networks 10(9), p. 1659.
[9] Thorpe S., Delorme, A., Van Rullen, R., 2001. Spike based strategies for rapid processing, Neural Networks 14(6-7), p. 715.
[10] Dayan, P., Abbot, L. F., 2005. Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems, MIT Press, Cambridge MA.
[11] Sutton, R. S., Barto, A. G., 1998. Reinforcement learning: an introduction. The MIT Press, Cambridge MA.
[12] Gu, Q., 2002. Neuromodulatory transmitter systems in the cortex and their role in cortical plasticity, Neuroscience 111, p. 815.
[13] Smith, W. B., Starck, S. R., Roberts, R. W., Schuman, E. M., 2005. Dopaminergic Stimulation of Local Protein Synthesis Enhances Surface Expression of GluR1 and Synaptic Transmission in Hippocampal Neurons, Neuron 45, p. 765.
[14] Florian, R. V., 2007. Reinforcement learning through modulation of spike-timing dependent synaptic plasticity, Neural Comput. 6, p. 1468.
[15] Izhikevich, E. M., 2003. Simple Model of Spiking Neurons, IEEE Trans. Neural Networks 14(6), p. 1569.
[16] Izhikevich, E. M., 2006. Polychronization: Computation with Spikes, Neural Computation 18, p. 245.
[17] Izhikevich, E. M., 2007. Solving the distal reward problem through linkage of STDP and dopamine signaling, Cereb Cortex 17, p. 2443.
[18] Braitenberg, V., Schütz, A., 1991. Anatomy of the Cortex, Springer-Verlag, Berlin.
[19] Abeles, M., 1991. Corticonics, Cambridge University Press, New York.